DOCUMENT RESUME

ED 330 724 TM 016 325

AUTHOR West, Sandra S.; Denton, Jon J.
TITLE An Empirical Validation of the Instrument: Student Perceptions of Teaching Effectiveness.
PUB DATE Apr 91
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Techniques; *Construct Validity; *Content Validity; Factor Analysis; Higher Education; Secondary Education; *Secondary School Students; Secondary School Teachers; State Programs; Student Attitudes; *Student Evaluation of Teacher Performance; Student Teachers; Teacher Certification; *Teacher Effectiveness; Teacher Interns; Test Validity
IDENTIFIERS Experts; *Student Perception of Teaching Effectiveness; *Texas Teacher Appraisal System

ABSTRACT

An assessment of the Student Perception of Teaching Effectiveness instrument (SPTE) is presented. The SPTE was developed to assess teaching interns' performance from the student's viewpoint. The Texas Teacher Appraisal System was selected as the content base for the development of the SPTE. The Texas system covers instructional strategies, classroom management and organization, presentation of subject matter, learning environment, and professional growth and responsibilities; the latter aspect was not addressed by the SPTE. This study sought to determine whether the SPTE provides useful information about the teacher's classroom effectiveness. The SPTE was administered to 480 secondary school students taught by 10 interns in an alternative teacher certification program. The SPTE's content validity was addressed through expert agreement; construct validity was explored with factor analysis. The SPTE was determined to exhibit content validity with respect to the state-required teacher observation system. However, the factor analysis yielded only a modest alignment of the empirical structure of the SPTE with the logically based structure of state-required teacher observation system. Ten tables, 3 figures, and a 62-item list of references are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED330724

An Empirical Validation of the Instrument: Student Perceptions of Teaching Effectiveness

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jon Denton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sandra S. West
Southwest Texas State University

Jon J. Denton
Texas A&M University

BEST COPY AVAILABLE

Paper Presented at the Annual Meeting
of the
American Education Research Association
Chicago, Illinois

April 6, 1991

TM016325

An Empirical Validation of the Instrument: Student Perceptions of Teaching Effectiveness

Abstract

This study was conducted to determine whether the Student Perception of Teaching Effectiveness instrument was valid. This instrument was developed to assess teaching interns' performance from the student's point-of-view. Validity of the instrument was addressed through the procedures, expert agreement and factor analysis for content validity and construct validity, respectively. The instrument was determined to exhibit content validity with respect to the state required teacher observation system. However, the construct validity component of the study yielded only a modest alignment of the empirical structure of the Student Perceptions Teaching Effectiveness instrument with the logically based structure of the state required teacher observation system.

The lack of public confidence in the American educational system during the recent past led to the accountability movement for teaching and teachers. The outcome of this movement resulted in a spate of methods for evaluating teaching performance. By the mid-eighties, at least 30 states had adopted systems for appraising teaching behavior (Sandefur, 1985). Similarly, Texas, through the enactment of the Texas Education Code, Section 13.302(b)(House Bill 72) required that a system be in place by the fall of 1986. The Texas Teacher Appraisal System (TTAS) became that system. It contains a component purporting to measure five domains of teaching effectiveness: instructional strategies, classroom management and organization, presentation of subject matter, learning environment, and professional growth and responsibilities. The TTAS is based on empirical evidence from the descriptive-correlational-experimental loop of process-product research which suggest that certain teaching behaviors, such as giving clear, explicit instructions increase student achievement (Brophy & Good, 1986; Dunkin & Biddle, 1974; Mitzel, 1960; Rosenshine & Furst, 1973; Shulman, 1986).

From the late 1950s through the early 1970s, a research strand evolved which concentrated on the development of classroom observation instruments and on descriptive, correlational, and experimental efforts. This effort became known as process-product research (Rosenshine & Furst, 1973). Processes here refer to those teaching processes that are effective in developing or increasing desired outcomes of the product (student achievement). Thus, process-product research has become the label to denote systematic exploration of teaching behaviors related to student outcomes using a series of related descriptive, correlational, and experimental studies (Gage, 1978).

Recent process-product research has been more comprehensive than earlier studies in identifying effective teaching practices because earlier work

relied on student achievement gains as the criterion of teaching performance. Brophy and Good (1986) have suggested that complex instructional problems require a flexible repertoire of effective responses. That is, the complex social setting called the classroom requires a multidimensional evaluation system. One dimension of such a system is the evaluation of effective teaching by students. Students represent a viable source of information regarding teaching behaviors that facilitate student learning in the classroom. Tucker (1979) pointed out that students are in the best position to see a normal lesson and can be considered to be capable consumers and observers. Further, a number of studies have supported the validity of student ratings for assessing teaching performance (Albrecht, Hanna & Hoyt, 1986; Chavez, 1984; Cohen, 1982; Draper, 1975; Fox, Peck, Blattstein, & Blattstein, 1983; McKelvey & Kyriacou, 1985).

While student ratings may provide valid data for assessing teaching performance, evidence is mixed regarding whether elementary school students can effectively evaluate the instructional behaviors of their teachers. However, extant studies indicate secondary students can assess accurately teaching performance. Moreover, substantial correlations between students' ratings and university and classroom supervisors' ratings of teachers' classroom performances have been reported. Generally, it appears that students' judgments about their teachers' behaviors agree with the judgments of experienced classroom observers.

Given the preceding findings on student assessment of teaching practices, student perception instruments which parallel official criteria could be used as part of a university or state evaluation system. Further, because some evidence suggests that student feedback can have more impact upon teaching improvement than the supervisor's feedback (Tuckman & Oliver, 1968), student

perceptions may be useful to supervisors and mentor teachers in helping teaching candidates to improve their teaching skills.

Operating from this frame of reference, the instrument, Student Perceptions of Teaching Effectiveness (SPTE), was developed to gather student perceptions of teaching candidates' instruction. This instrument was designed to be congruent with the state mandated TTAS.

Purpose and Research Questions

The primary purpose of this study was to validate the Student Perception of Teaching Effectiveness (SPTE) instrument. This instrument, used to evaluate a teaching candidate's performance from the student's point of view, closely adheres to the indicators in the criteria and domains of the Texas Teacher Appraisal System. (It is important to note that the Texas teacher evaluation system contains no component which gather the students' assessments of their teachers' behaviors.)

The following research questions were posed to examine the issue of whether an instrument (SPTE), designed to gather student perceptions of a teacher's instructional effectiveness, provides worthwhile information about the teacher's classroom effectiveness.

1. Does the Student Perception of Teaching Effectiveness instrument have content validity with respect to the Observation Record/Evaluation Record of the Texas Teacher Appraisal System?
2. Are the logical clusters of items on the Observation Record/Evaluation Record of the Texas Teacher Appraisal System compatible with empirically derived item clusters of the Student Perception of Teaching Effectiveness instrument?

Related Literature

Multiple Data Sources

The evaluation of the effectiveness of certain teaching behaviors is sufficiently controversial that many recommend using multiple data sources in rendering judgments about a teacher's effectiveness. (Barrett, 1986; Bonfadini, 1985; Capie, Johnson, Anderson, Ellett, & Okey, 1979; Florida Coalition for the Development of a Performance Measurement System, 1983; Kauchak, Peterson, & Driscoll, 1984; Larson, 1984; Manatt, 1987; McDonald, 1979; McGreal, 1983; McKelvey & Kyriacou, 1985; Meighan, 1978; Peterson, K., 1987; Peterson, P., 1988; Peterson & Kauchak, 1982, Soar, Medley, & Coker, 1983). The most prevalent justification given by these authors is that the complexity of teaching requires extensive data sources and indicators of quality. One complication of identifying data sources of teaching effectiveness is that teachers are not entirely responsible for even the immediate outcome of their efforts or talents. Much of what teachers do is context dependent; what is effective in one place with one kind of student is not good practice in another setting with another kind of learner. Other areas such as school climate, student effort, prior knowledge, as well as home and community environment affect the amount learned by the student. Finally, only some goals of teaching are visible in the short term and are easily measured. Thus, it is important when evaluating teaching behavior that a variety of teaching quality assessments be used to balance and weigh the preceding factors according to the goals of the evaluation.

Our current knowledge suggests that there is no single procedure or single criterion that can be depended upon to identify the effective teacher. Gage (1983) stated, "Research on teacher effect provides suggestions about

how a teacher should behave on a continuum from a hunch to a trace to an imperative to a categorical imperative" (p. 496). Manatt (1987) further adds, "Some of the evidence is so strong that for a teacher to admit certain behaviors probably constitutes malpractice" (P. 11). However, conclusive evidence does not exist that teachers are responsible in a causative sense for a major portion of a student's achievement, yet the heart of accountability rests on such linkage (McDonald, 1979). Effectiveness is context bound. It is with all of these factors in mind that McDonald (1979) stated that great humility should be used in making judgments about teacher effectiveness from the kinds of knowledge and skill presently possessed, and therefore, common sense dictates the use of multiple techniques and multiple sources.

In a multiple data source system that is competency based, a specific competency can be assigned to a single data source(i.e., classroom management to an observation scale, planning to the portfolio) or to multiple data (i.e., interactive patterns, class observations, student questionnaires, principal ratings). In many systems, each competency is assigned to at least two data sources. This latter approach provides for an assessment of the desired teaching performance that is less affected by the peculiarities of a specific respondent or data collection method. Additionally, it compensates for the teacher who may have great difficulty in conducting a class when an evaluator is present. Thus, the multiple data source system provides counterbalances to offset variance associated with varying performances and characteristics of particular data collection methods.

Student Perceptions in a Multiple Data Source System

Past investigations, suggest that pre-college students have reliable and valid perceptions of effective teaching performance. Students can evaluate teaching behaviors in a reliable and consistent manner (Christensen, 1960;

Clark, 1987; Denton, Calarco & Johnson, 1977; Hanna, Aubrecht, & Hoyt, 1983; Lawrenz, 1977; Martin, 1987; Masters, 1978; Thompson, 1975; Veldman & Peck, 1967; Waxman & Duschl, 1987). McKelvey and Kyriacou (1985) have reported that pupil ratings of teaching behaviors are valid in terms of being stable over time, and in agreement with conventional evaluators such as other teachers, teacher educators and/or administrators.

In an earlier effort, Veldman and Peck (1967) surveyed secondary students of 554 student teachers using the Pupil Observation Survey and found secondary students' perception of teaching performance to be reliable and valid. They suggested that pupil ratings have the advantage of averaging a large number of individual biases. Additionally, these researchers note, student evaluations are the product of observing the teacher on many occasions under normal conditions.

The effect that student feedback has on altering teaching behavior has been examined as well (Lacy, Tobin & Treagust, 1984; Lewis & Bartholomew, 1984; Tuckman & Oliver, 1968; Waxman & Duschl, 1987). Results suggest that student feedback is a viable and effective mechanism for providing teachers with information about their classroom behavior. This feedback can create an imbalance in teachers' perceptions of their own behavior. Posner, Strike, Hewson, and Gertzog (1982) called this a "conceptual change" in the teacher's central commitments. This conceptual change or imbalance exists whenever teachers discover that their attitudes to or perceptions about their teaching differ from that of their students. Gage (1972) has suggested that this state of imbalance motivates teachers to do something about their behavior in order to restore themselves to a balanced condition. Further evidence of the impact of student ratings on teachers is the finding by Evertson, Anderson, Anderson, and

Brophy (1980) that significant correlations exist between effective teachers and their receptivity to student input.

In summary, research literature and professional experience suggest that student reports and evaluations of teachers, particularly in reference to discrete and visible behaviors, are potentially important sources of information for teacher evaluation and self-improvement. It is hoped this instrument and resultant data will extend this literature.

Methods

Instrument Development

The Texas Teacher Appraisal System (1986) was selected as the content base for the development of the SPTE instrument for two reasons. First, the TTAS is the legal instrument for assessing teachers in Texas; and second, the TTAS contains behavioral indicators of teachers that have been linked with effective instruction.

The TTAS consists of two parts: (a) Teacher Assessment of Instructional Goals and Outcomes and (b) Observation Record/Evaluation Record (OR/ER). Although the states of Georgia and Texas refer to their entire system as an evaluation instrument, in reality only certain components are used for teacher evaluation. In Georgia, only the Teaching Plans and Materials, Classroom Procedures, and Interpersonal Skills sections are used for certification purposes (Georgia Department of Education, 1984). The Texas system uses the Observation Record/Evaluation Record (OR/ER) as the teacher observation/evaluation instrument.

Based on the models of evaluation established by states such as Florida and Georgia, the OR/ER includes several research-validated domains of teaching behaviors. The Texas Education Agency Teacher Orientation Manual (1988) defines the Texas instrument as a "list of 65 specific teaching behaviors

(performance indicators) categorized into 13 subsets called criteria; these criteria are grouped into five major areas called domains; used in the OR/ER to assess teacher performance" (p.112). The behavioral descriptions for each indicator are consistent with their use in the teacher effectiveness literature and serve two purposes: first, to communicate the instrument's meaning of the behavior as derived from research; and second, to establish ground rules for identifying and assessing behaviors during classroom observations.

Figure 1 provides an excerpt from the OR/ER which illustrates the domain, the criterion, and the indicators. The words in parentheses are added by the authors for clarity.

(Domain)	I.	Instructional Strategies
(Criterion)	1.	Provides opportunities for student to participate actively and successfully.
(Indicators)	a.	varies activities appropriately
	b.	interacts with group(s)
	c.	solicits student participation

FIGURE 1. Excerpt from the OR/ER.

The five domains in the OR/ER include: I. Instructional Strategies; II. Classroom Management and Organization; III. Presentation of Subject Matter; IV. Learning Environment; and V. Professional growth and Responsibilities. Indicators from the initial four domains of this instrument provided direction and content for the development of the instrument called the Student Perception of Teaching Effectiveness instrument (SPTE). SPTE items were not written for the 16 indicators of Domain V because it was assumed that students do not typically think of their teachers in terms of assessing those qualities associated with the professional development of teachers.

Each of the remaining 49 indicators from Domains I-IV in the OR/ER was reviewed, and one or more corresponding behaviorally-based statements were written applying the following criteria: student's viewpoint, vocabulary, readability, and understanding. All SPTE statements were stated positively, in keeping with the positive wording of the OR/ER indicators. Additionally, the need for brevity of the total instrument was considered important for the ease of administration as well as the issue of student willingness to complete a brief survey. Figure 2 contains a sample of three OR/ER indicators and the corresponding SPTE statements.

Or/ER Indicator	Corresponding SPTE Statement
"Solicits student participation"	"The teacher encourages students to participate in class rather than just listen."
"Implements at appropriate level"	"The teacher instructs the class at an appropriate level, not too difficult and not too easy."
"Communicates learning expectations"	"The teacher communicates to students what they are to learn as a result of the lesson."

FIGURE 2 Sample of OR/ER indicators and corresponding SPTE statements.

The SPTE was designed to be administered to students in grades 8 through 12. It was felt the reading level as well as the vocabulary needed to be appropriate for secondary students. A subsequent check of the reading level of the SPTE using the Fry Readability procedure yielded the readability level of grade seven.

Comprehension was also considered an important factor because some of the indicators in the OR/ER were written in technical educational language

that was considered to be abstruse for some students. Those items were rewritten using vocabulary and phrases that would make the teaching behavior being considered understandable to secondary students.

As noted previously, indicators from Domain V, Professional Growth and Responsibilities, were not transformed into SPTE statements because students have few opportunities to develop perceptions of teachers' professional growth and responsibilities. No items were written for four of the indicators because use of these indicators would have resulted in redundant items from the students' perspective, and three other indicators were omitted because they appeared difficult for students to assess from the frame of reference of a student in the class. Table 1 provides a summary of the consolidated indicators and the resulting SPTE statements and omitted OR/ER indicators.

TABLE 1

Summary of Consolidated OR/ER Indicators/SPTE Items and the Omitted OR/ER Indicators

1. Consolidated OR/ER Indicators		SPTE Item
2b.	"Monitors student performance"	"The teacher asks student to demonstrate they know the material in a lesson"
2c.	"Solicits responses/demonstrations for assessment"	
2d.	"Reinforces correct response/ performances."	"The teacher tells students when their answers are correct and complete."
2e.	"Provides corrective feedback/ clarifies/none needed."	
4e.	"Keeps students engaged."	"The teacher identified students who are doing something other than the assigned task and guides them back to their classwork."
5b.	"Prevents off-tasks behavior/ none needed	
5c.	"Redirects/stops inappropriate/ disruptive behavior/none needed."	
Omitted Indicators		
1b.	"Interacts with groups appropriately."	No related item
3d.	"Maintaining appropriate seating/ grouping"	No related item
6b.	"Presents information in appropriate sequence"	No related item
Domain V		No related items

Seven of the 49 indicators of effective teaching behaviors were not used to develop SPTE statements resulting in 42 of the indicators being matched with at least one corresponding SPTE item. Four indicators served as the basis for two SPTE items per indicator because other influences were considered and

additional emphases were placed on those indicators. These decisions yielded a 46 item instrument.

Students were requested to mark their answers anonymously on optical scan sheets. Additionally, the format of the scale was adjusted to permit students to respond to each statement with respect to their degree of agreement or disagreement. The responses were recorded on an optical scan sheet with a five-point Likert scale: Strongly Agree (5), Agree (4), Undecided (3), Disagree (2), and Strongly Disagree (1). The Likert scale was chosen for its ease of use and established validity.

Upon completion by the students, the optical scan sheets were submitted to the teaching interns. All questionnaires and answer sheets were returned to the university supervisor and subsequently scored at the Test and Measurement Center at the university.

Sample

The sample was comprised of secondary students of ten teaching interns who were enrolled in an alternate certification program. Four hundred eighty students completed the SPTE in May during the final two weeks of the school year. The following protocol was followed in gathering and analyzing the student generated data.

Protocol for Administering SPTE

1. A packet of 50-75 instruments and 100-150 optical scan sheets was provided for each intern. Additional optical scan sheets were provided to enable the interns to administer the instrument to all of their class sections.

2. Students were requested to record their answers anonymously on the optical scan sheets. This protocol encouraged the students to be honest in recording their perceptions without the threat of retribution from the intern teacher.
3. The interns returned the optical scan sheets to the project coordinator one week after administration of the SPTE. This deadline was established to ensure that the optical scan sheets would not be misplaced in the hectic days at the conclusion of the school year.
4. All optical scan sheets were examined to determine whether they were marked with pencil, whether any items were marked with multiple responses, and whether item responses created graphics patterns. If problems were evident, the sheet was discarded.
5. The optical scan sheets were machine scored at the university's Test and Measurement Center.
6. Data files were stored on a mainframe computer at the university for subsequent data analysis.
7. SPTE item means were shared with each teaching intern.

Content Validation

Since the SPTE items were derived from the OR/ER indicators, an effort was undertaken to verify content validity, i.e., the degree to which the items represented the content the items purported to represent. The SPTE was designed to reflect the OR/ER indicators of effective teacher behaviors. The indicators were stated in the OR/ER and clarifying statements were available in the TTAS training manual which defined in observable terms the specific effective teacher behaviors sampled by the SPTE.

The SPTE was submitted to a panel of eight experts during the spring of 1989. The panel served as judges to determine the content validity of the instrument. Experts in this case were defined in terms of the individual's training and experience with the TTAS. Two of the judges were Educational Service Center TTAS trainers and six were school administrators who were active in evaluating teachers in their respective school districts using the TTAS. Each panel member received a packet of material which included a letter of request with instructions, a specially marked SPTE, and a copy of the OR/ER. Figure 3 provides a sample of the specially marked SPTE from the expert panel's packet.

The letter of request containing the instructions asked each person to agree or disagree with the correspondence between the criterion indicator in the OR/ER and the indicated item of the SPTE. In order to obtain a 100 percent return, a follow-up request was made to those panelists whose packets were not returned within two weeks.

A decision-rule regarding agreement of the judges was made prior to receiving completed instruments from the panel members. The criterion level was set at .80, i.e., if at least 80 percent of the judges agreed that the item on the SPTE corresponded with the OR/ER indicator, then it was assumed that content validity of that item had been demonstrated.

STUDENT PERCEPTIONS OF TEACHING EFFECTIVENESS

				Agree	Disagree
Domain 1 Criterion 1 Indicator	c	1.	The teacher encourages students to participate in class rather than just listen.	___	___
	a	2.	The teacher instructs using more than one approach, that is, large group, small group and to individuals.	___	___
	d	3.	The teacher encourages and calls on non-volunteers to participate in class.	___	___
	e	4.	The teacher instructs the class at an appropriate level, not too difficult and not too easy.	___	___

FIGURE 3. Sample of the specially marked SPTE.

Comparison of Logical - Empirical Item Clusters

A factor analysis was conducted on the SPTE data to identify the groups of variables that were moderately or highly correlated with one another so as to determine whether the logical clusters of indicators in the OR/ER were compatible with empirically derived factor analyzed clusters of the SPTE. The unit of analysis for the factor analytic procedures was the student. The factor analysis subprogram from SPSSx was used to conduct the analysis. The compatibility of the logical clusters on the OR/ER and the empirically derived clusters of the SPTE was then determined.

Results

Research question number one addressed the issue of the content validity of the instrument, Student Perceptions of Teaching Effectiveness. The SPTE was sent to a panel of eight judges to determine content validity of the items. All judges rated and returned the material. Percentages of agreement

were calculated for each of the items regarding the judges' agreement with the correspondence between each criterion indicator in the OR/ER and the indicated item of the SPTE. Table 2 presents the percentage of agreement of each of the items.

TABLE 2

Percentage of Judge Agreement of SPTE Items With Respect to the OR/ER Indicators*

Item	Percent	Item	Percent	Item	Percent
1	75	17	100	32	100
2	100	18	100	33	63
3	75	19	88	34	100
4	100	20	100	35	100
5	100	21	100	36	63
6	100	22	100	37	100
7	100	23	100	38	100
8	75	24	100	39	100
9	100	25	75	40	100
10	100	26	88	41	100
11	100	27	100	42	100
12	100	28	100	43	100
13	100	29	100	44	63
14	100	30	75	45	100
15	100	31	100	46	100
16	100				

*Items rounded to nearest percentage

The criterion level for acceptance of the listed SPTE items was set at .80 prior to receiving responses from the judges. Thirty-eight of the 46 items reached this level regarding their acceptance as reflecting the corresponding indicator in the OR/ER. The panel agreed 100 percent with 36 items. Agreement levels ranging between 80 percent and 100 percent were found for

38 of the items. Eight items did not reach the criterion level for acceptance. The level of agreement was 75 percent on five items and 63 percent on three items.

All items were regarded by the judges as reflecting at least one indicator in the OR/ER. Although changes were suggested for assigning different indicators to a few items, the substance of the SPTE was judged to be a valid content representation of the teaching behaviors represented on the OR/ER.

Research question two explored whether logical clusters of the indicators on the OR/ER were compatible with the empirically derived clusters of the items on the SPTE. Discussion of the suggested criteria for choosing the "best" number of factors is presented in the following paragraph prior to presenting the results of the factor models.

The two criteria used to select the model were obtained from suggestions offered by Rummel (1970) and Hair, Anderson, and Tatham (1987), that is, (a) statistical indicators for selecting the "best" number of factors and (b) an analysis of the content of the factors. Selection rules for the best number of factors have been developed for eigenvalues, scree plots, percent of variance accounted for by the model, overlap in factor loadings, and loading values. The selection rules include: a factor must have an eigenvalue of one or greater to be considered significant. Second, a scree plot indicates the maximum number of factors to extract when the plot becomes horizontal, that is, when the curve first beings to straighten out. As a general rule, the scree tail test will result in at least one more factor being considered significant than will the latent root criterion (Hair et al., 1987). Third, the percent of variance accounted for should be as great as possible in considering the best number of factors. Fourth, the choice with the best number of factors should have the least amount of overlap in item factor loadings. Finally, the loading values with the largest absolute factor loading is the optimal choice. The loading values should be at least .30

to be considered significant, while factor loadings of .50 or greater are considered very significant. Ultimately, the number of significant loadings on each column of the factor matrix or loading associated with one variable, needs to be maximized. And the number of loadings with negligible values need to be minimized (Hair et al, 1987). Thus, a dually loaded item would be placed in the factor with the higher loading. These recommendations guided the decisions occurring throughout the analysis associated with research question two.

Factor analysis of the SPTE was conducted with the factor analysis subprogram from SPSSx (1986). Principal-component analysis was performed on the initial set of data (N=480) from students of interns. Three factors were extracted which had eigenvalues of greater than one (see Table 3).

TABLE 3

Final Statistics of Principal Component Analysis

Factor	Eigenvalue	% of var	Cum %
1	16.3	35.5	35.5
2	1.5	3.2	38.7
3	1.0	2.2	40.8
4	.82	1.8	42.6
5	.72	1.6	44.2
6	.68	1.5	45.6
7	.54	1.2	46.8
8	.49	1.1.	47.9

The initial extraction was followed by an extraction utilizing principal factoring with iteration. Additionally, both oblique (oblimin) and orthogonal (varimax) rotations were requested. The oblique rotation failed to converge in 25 iterations. The orthogonal rotation extracted a model with eight factors. The

eight factor model was rejected because eigenvalues did not reach the criterion (Eigenvalue 1) for five of the factors. Further, the scree plot curve first became horizontal at factor four. Given the conceptual base of the four domains of the TTAS, it was expected that a four factor model would meet the afore mentioned criteria. However, the empirical model suggested a different ordering.

A three factor model was extracted to determine whether three factors presented a better choice considering the criteria associated with eigenvalues, scree plots, percent of explained variance, overlap in factor loadings, magnitude of loading values and content analysis. The final statistics of the factor analysis for the three factor model revealed that two factors had eigenvalues of one or greater as shown in Table 4. Subsequently, an extraction of a two factor model was performed in order to determine the optimal empirical model.

TABLE 4

Final Statistics of Factor Analysis for Three Factor Model

Factor	Eigenvalue	% of var	Cum %
1	16.3	35.3	35.3
2	1.4	3.0	38.3
3	.92	2.0	40.3

The final statistics of the two factor model, as shown in Table 5 produced two factors which had eigenvalues of one or greater.

TABLE 5

Final Statistics of Factor Analysis of the Two Factor Model

Factor	Eigenvalue	% of var	Cum %
1	16.2	35.3	35.3
2	1.4	3.0	38.3

As Table 6 reveals, 39 of the items loaded in factor 1 and 35 of the items in factor 2. Factor 1 contained 11 items which loaded exclusively on this factor with loadings ranging from .46 to .67. Seven items loading exclusively on factor 2 yielded loadings ranging from .38 to .62

TABLE 6

Two Factor Model: Factor Loadings for STPE Items*

Item	Factor 1	Factor 2
17	.67	
16	.65	.31
37	.64	
6	.62	
26	.62	
9	.62	
5	.62	
31	.61	.33
45	.59	.36
39	.59	.32
2	.56	
4	.55	
12	.55	.42
38	.54	
44	.54	.38
22	.54	.31
29	.54	.35
28	.53	.35

TABLE 6

Continued

Item	Factor 1	Factor 2
25	.55	.42
32	.52	.38
24	.52	
1	.51	.35
40	.47	.45
21	.47	.43
41	.46	
27	.46	.34
33	.45	.44
7	.41	.35
8	.35	.35
19		.62
35	.35	.61
34	.40	.54
20		.54
10	.33	.53
43	.33	.52
23	.35	.52
36	.42	.52
42	.44	.51
13		.47
14		.46
18		.45
30	.37	.45
11	.32	.45
46	.37	.43
3		.42
15		.38

*Item loadings rounded to the nearest one-hundredth.

In order to analyze how the theoretical construct (i.e., domains of the OR/ER on which the SPTE was based corresponded to the two factor empirical structure, the following analysis was conducted. Items in each factor were classified with respect to the OR/ER domain from which they were derived. Given the criteria cited previously, items were assigned to the domain which

had the higher loading value. The results of these classifications are presented in Table 7.

Items clustered in factor 1 related primarily to instruction. fifty-eight percent of the items in this factor were from a combination of Domains I and III of the OR/ER (instructional strategies and presenting subject matter). Yet 42 percent of the items loaded heavily in the remaining two domains. For example, item 17, which had a loading of .67, stated, "The teacher maintains proper pace with the lessons, that is, the class does not drag or seem to be hurried." This item on pacing is found in Domain II (classroom management) in the OR/ER.

Forty-seven percent of the factor 2 items were from Domain II, classroom management. Items loading exclusively in this factor [i.e, items 19 (.62) and 20 (.54)] with the largest loadings belonged in Domain II. Table 7 provides a listing of item numbers and the OR/ER domain referent found in each factor.

TABLE 7

Construct Validity Assessment of Two Factor Model with Domains of OB/ER

Item	Factor 1 OB/ER domain	Factor 2
17	II	
16	II	
37	IV	
6	I	
26	III	
9	I	
5	I	
31	III	
45	IV	
39	IV	
2	I	
4	I	
12	II	
38	IV	
44	IV	
22	II	
29	III	
28	III	
25	III	
32	III	
24	III	
1	I	
40	IV	
21	II	
41	IV	
27	III	
33	III	
7	I	
8	I	
19		II
35		III
34		III
20		II
10		II
43		IV
23		III
36		III
42		IV
13		II
14		II
18		II
30		III
11		II
46		IV
3		I
15		II

Given the criteria linked with eigenvalues, screeplots, percent of variance, loading values, overlap of item loadings and empirical content analysis attributed to Rummel (1970) and Hair et al. (1987) a clear distinction

was expected. However, these criteria did not provide a model whose construct of effective teaching behaviors could be clearly discriminated. Consequently, a decision rule was adopted that only items which loaded exclusively on a single factor would be considered as part of the factor to determine the content structure of the factors.

Using this decision rule, 64 percent (7 out of 11) of the items in factor 1 came from Domains I and III (instructional strategies and presentation of subject matter) of the OR/ER. In addition, three items (27 percent) came from Domain IV.

Factor 2 yielded seven items which loaded exclusively on one factor. Six out of seven (86 per cent) of items were from Domain II (classroom management). Table 8 presents the items which loaded only in one factor.

TABLE 8

Construct Validity Assessment of Two Factor Model with Domains of OR/ER of Singlely Loaded Items

Item	Factor 1 OR/ER domain	Factor 2 OR/ER domain
17	II	
37	IV	
6	I	
26	III	
9	I	
5	I	
2	I	
4	I	
38	IV	
24	III	
41	IV	
19		II
20		II
13		II
14		II
18		II
3		I
15		II

In summary, given the decision rule to use only items that loaded on a single factor, factor 1 consisted primarily of items relating to instruction and items constituting classroom management were found in factor 2.

Discussion

The first research question focused on the issue of the Student Perception of Teaching Effectiveness (SPTE) instrument's content validity. Results suggest that the SPTE is characterized by a high degree of content validity. Ratings to 36 of the 46 items reached the established criterion level of .80. Further, all items were determined by the panel of judges to have content validity with respect to the Observation Record/Evaluation Record (OR/ER) of the Texas Teacher Appraisal System (TTAS).

This finding supports the idea that behavioral indicators of effective teachers identified from the process-product research paradigm (Brophy & Evertson, 1974; 1976; Evertson, Anderson, Anderson & Brophy, 1980; Evertson & Emmer, 1982; Sanford & Evertson, 1980) can be integrated into alternate forms for evaluating classroom practices of secondary level teachers. For this investigation, the behavioral indicators of effective teachers were crafted into statements that secondary students could understand in evaluating the instructional effectiveness of their teacher. The issue being examined with respect to this research question was whether the SPTE statements accurately represented the OR/ER indicators on which they were based. Teacher evaluators skilled in using the OR/ER agreed the SPTE statements do relate to the OR/ER indicators.

Research question two addressed the construct validity of the SPTE through examining the compatibility of the logical clusters in the OR/ER and the empirically derived clusters in the SPTE. As suggested by Rummel (1970) and Hair et al. (1987), the two major criteria to be used to select the optimal model

are (a) criteria for the "best" number of factors, and (b) an analysis of the content of the factors.

The criteria for the optimal number of factors include eigenvalues, scree plots, percent variance, overlap in factor loadings and loading values. As noted in the results, eigenvalues reached the criterion value of 1 for three factors on the principal component analysis. This finding was unexpected, given the conceptual base of the 4 domains of the TTAS. The scree plots became horizontal at factor four. The percent of variance accounted for by the first three factors was 40.8. Subsequently, the extraction of a three factor model revealed that only two factors had eigenvalues of 1 or greater. Finally, a two factor model was extracted to determine the optimal empirical model. This extraction produced two factors which had eigenvalues of 1 or greater that accounted for 38.3 percent of the variance.

Factor one loaded primarily with items which addressed instruction. This finding is consistent with the findings on certain effective teaching behaviors. For example, an effective teacher provides opportunities for students to participate actively and successfully. In particular, students learn more efficiently if they are engaged in activities where instruction is implemented at the appropriate level of difficulty and otherwise suited to their current achievement levels and needs (Okey, Capie, Ellett & Johnson, 1978; Stallings, 1976). Item 5 addressed this particular instructional strategy by stating, "The teacher instructs the class at an appropriate level, not too difficult and not too easy." Similarly, item 4, which asked for information concerning whether the teacher provides ample time for students to respond to teacher questions of solicitations loaded on factor one. This, too, is consistent with findings on wait time (Rowe, 1974).

Items which loaded only in factor one are presented in Table 9. All of the items except items 38, 37, and 41 occur in Domains I (instructional strategies) and III of the OR/ER. Although items 38 and 37 are located in Domain IV (learning environment) in the OR/ER, it is plausible for these items to load in the factor on instruction. Items 38 and 37 stated respectively, "The teacher emphasized the value of the content or lesson to students," and "The teacher relates class content to studer't interests." The placement of these items in the factor on instruction seemed appropriate given the research on learning which reports the use of materials that relate content to student prior knowledge and stating why the content of the lesson is important positively affects student achievement (Capie, Tobin & Bowell, 1980; Okey et al, 1978). However, the content of item 41, "The teacher avoids sarcasm and negative criticism to students," from the learning environment domain, while clearly related to teacher comments in the classroom, appears to match the logical descriptor of learning environment more appropriately than instruction.

TABLE 9

SPTE Items in Factor 1 in Two Factor Model

17.	The teacher maintains proper pace with the lessons, that is, the lessons do not drag or seem to be hurried.
37.	The teacher relates class content to student interests.
6.	The teacher communicates to students what they are to learn as a result of the lesson.
9.	The teacher reteaches a topic a different way if the class does not understand.
5.	The teacher instructs the class at an appropriate level, not too difficult and not too easy.
2.	The teacher instructs using more than one approach, that is, large group, small group and to individuals.
4.	The teacher provides time for students to respond to instructional questions.
38.	The teacher emphasizes the value of the content or lesson to students.
24.	The teacher relates the content of the lesson to previous lessons or experience.
41.	The teacher avoids sarcasm and negative criticism to students.

Factor two, classroom management, consisted of items such as 19, "The teacher explains expectations for class behavior" and 13, "The teacher has materials, equipment, and the laboratory ready when class begins." The importance of classroom management has been clearly linked to student achievement (Brophy & Evertson, 1974; Evertson, 1985; Evertson & Weade, 1989; Rupley & Blair, 1987).

Items 19, 20, 13, 14, 18, 3, and 15 loaded solely in factor two. All of the items except item 3 were found in Domain II, classroom management. Item 3 stated, "The teacher encourages and calls on non-volunteers to participate in class." This effective teaching behavior could be interpreted as playing a role in classroom management as well as instruction. Table 10 lists the items that loaded exclusively in factor two.

TABLE 10

SPTE Items in Factor 2 in the Two Factor Model

19.	The teacher explains expectations for class behavior.
20.	The teacher identifies students who are doing something other than the assigned task and guides them back to their classwork.
13.	The teacher has materials, equipment, and the laboratory ready when class begins.
14.	The teacher always begins class promptly.
18.	The teacher stays on the topic.
3.	The teacher encourages and calls on non-volunteers to participate in class.
15.	The teacher avoids wasting time at the end of the period.

An uncritical acceptance of the logical basis for the four domains of the TTAS led to an expectation that a four factor empirical model would result from these analyses. However, the factor analysis procedures yielded an empirical model with two factors. Using the decision rule to retain only those items that

loaded on a single factor, and an analysis of the content of those items, factor 1 was labeled instruction while factor 2 was labeled classroom management. These factor names appeared to be the optimal conceptual and empirical choices given the results of this investigation. In closing, a caution in determining construct validity of a perception scale using factor analysis procedures is to NOT assume compatibility between logically and empirically derived structures of the instrument.

REFERENCES

Albrecht, J., Hanna, G., & Hoyt, D. (1986). A comparison of high school student ratings of teaching effectiveness with teacher self-ratings: Factor analytic and multitrait-multimethod analyses. Educational and Psychological Measurement, 46 223-231.

Barrett, J. (1986). The evaluation of teachers. (ERIC Digest 12 and ERIC Clearinghouse on Teacher Education. ED No. 278657).

Bonfadini, J. (1985). Improving student-teacher performance. Technology-Teacher, 44(6), 8-10.

Brophy, J., & Evertson, C. (1974). Process-product correlations in the Texas Teacher Effectiveness Study: Final Report, No. 74-4 Austin, TX: The University of Texas, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 091 394).

Brophy, J., & Evertson, C. (1976). Learning from teaching: A developmental perspective. Boston: Allyn and Bacon.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.). Handbook of research on teaching: Third edition. New York: Macmillan.

Capie, W., Johnson, C., Anderson, S.J., Ellett, C., & Okey, J. (1979). Teacher Performance Assessment Instruments. Athens, GA: Teacher Assessment Report, University of Georgia.

Capie, W., Tobin, K.G., & Bowell, M. (1980). Using science achievement to validate ratings of student-teacher competencies. Paper presented at the National Association for Research in Science Teaching, Boston. (ERIC Document Reproduction Services No. ED 186-261).

Chavez, R.C. (1984). The use of high-inference measures to study classroom climate: A Review. Review of Educational Research, 54, 237-261.

Christensen, C. (1960). Relationship between pupil achievement, pupil affect need, teacher warmth, and teacher permissiveness. Journal of Educational Psychology, 51, 169-174.

Clark, D.L. (1987). High school seniors react to their teachers and their schools. Phi Delta Kappan, 68(7), 503-509.

Cohen, P.A. (1982). Synthesizing research results on teacher evaluation using meta-analytic procedures. Hanover, NH: Dartmouth College. (ERIC Document Reproduction Service No. ED 223 646).

Denton, J.J., Calarco, J., & Johnson, C. (1977). Pupil perceptions of a student teacher's competence. Journal of Educational Research, 70, 180-185.

Draper, T. (1975). Children talking about their mathematics teachers. Mathematics Teaching, 72, 32-35.

Dunkin, J.J., & Biddle, B.J. (1974). The study of teaching. New York: Holt, Rinehart and Winston.

Evertson, C.M. (1985). Differences in instructional activities in high and low achieving junior high classes. Austin: The University of Texas, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 195 546).

Evertson, C.M., Anderson, C., Anderson, L.M., and Brophy, J. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. American Educational Research Journal, 17, 43-60.

Evertson, C.M., & Emmer, E. (1982). Effective management at the beginning of the school year in junior high classes. Journal of Educational Psychology, 74(4), 485-498.

Evertson, C.M., & Weade, R. (1989). Classroom management and teaching style: Instructional stability and variability in two junior high English classrooms. The Elementary School Journal, 89(3), 379-392.

Florida Coalition for the Development of a Performance Measurement System. (1983). Florida Performance Measurement System. Chipley, FL: Panhandle Area Educational Cooperative.

Fox, R., Peck, R.F., Blattstein, A., & Blattstein, D. (1983). Student evaluation of teacher as a measure of teacher behavior and teacher impact on students. Journal of Educational Research, 77, 16-21.

Gage, N. (1972). Teacher effectiveness and teacher education. Palo Alto, CA: Pacific.

Gage, N. (1978). The scientific basis of the art of teaching. New York: Teachers College Press, Columbia University.

Gage, N. (1983). When does research on teaching yield implications for practice? Elementary School Journal, 83, 492-496.

Georgia Department of Education. (1984). Teacher performance assessment instruments. Atlanta, Georgia Department of Education, Division of Staff Development.

Hair, J., Anderson, R., and Tatham, R. (1987). Multivariate Data Analysis (2nd ed.). New York: Macmillan.

Hanna, G.S., Aubrecht, J.D., & Hoyt, D.P. (1983). Discriminant and convergent validity of high school student ratings of instruction. Educational and Psychological Measurement, 43, 873-878.

Kauchak, D., Peterson, K., & Driscoll, A. (1984). An interview study of teachers' attitudes toward teacher evaluation practices. A paper presented at the annual meeting of the American Education Research Association. (ERIC Document Reproduction Service ED 246 106).

Lacy, T., Tobin, K.G., & Treagust, D.F. (1984). Development, validation and reliability of the elementary science learning environment inventory. Paper presented at the Annual Conference of International Association for Student Assessment, Perth,Australia.

Larson, R. (1984). Teacher performance evaluation-what are the key elements? NAASP Bulletin, 68(469), 13-18.

Lawrenz, F. (1977). The stability of student perception of the classroom learning environment. Journal of Research in Science Teaching, 14(1), 77-81.

Lewis, M., & Bartholomew, R. (1984). The effectiveness of pupil feedback and self-analysis in changing selected science teaching practices. In L.J. Bethel (Ed.), Research and curriculum development in science education. Austin: Science Education Center, University of Texas.

Manatt, R.P. (1987). Lessons from a comprehensive performance appraisal project. Educational Leadership, 44(7), 8-14.

Martin, D.V. (1987). Use of stimulated recall in video tape analysis of student teacher development. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Arlington, VA (ERIC Document Reproduction Services No. ED 277 700).

Masters, J.R. (1978). High school student ratings of teachers and teaching methods. Journal of Educational Research, 72, 219-224.

McDonald, J.B. (1979). Planning for the evaluation of teaching. In W.R. Duckett (Ed.), Planning for the evaluation of teaching. Bloomington, IN: Phi Delta Kappa.

McGreal, T.L. (1983). Successful teacher evaluation. Alexandria,VA: Association for Supervision and Curriculum Development.

McKelvey, J., & Kyriacou, C. (1985). Research on pupils as teacher evaluators. Educational Studies, 11(1), 25-31.

Meighan, R. (1978). A pupil's eyeview of teaching performance. Educational Review, 30(2), 125-137.

Mitzel, H.E. (1960). Teacher effectiveness. In C.W. Harris (Ed.). Encyclopedia of Educational Research (3rd ed.), 1481-1486. New York: Macmillan.

Okey, J., Capie, W., Ellett, C.D., & Johnson, C.E. (1978). Teacher performance validation studies. Paper presented at the annual meeting of the Georgia Educational Research Association, Atlanta.

Peterson, K. (1987). Teacher evaluation with multiple and variable lines of evidence. American Educational Research Journal, 24(2), 311-317.

Peterson, P.L. (1988). Teachers' and students' cognitional knowledge for classroom teaching and learning. Educational Researcher, 17(5), 5-14.

Peterson, K., & Kauchak, D. (1982). Teacher evaluation: Perspectives, practices, and promises. Salt Lake City: Utah Teacher Evaluation Project, Center for Educational Practice. (ERIC Document Reproduction Services No. ED 233 996).

Posner, G., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodation of a science conception: Toward a theory of conceptual change. Science Education, 66, 527-540.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.). Second handbook of research on teaching. Chicago: Rand McNally.

Rowe, M.B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic, and fate control: Part one, wait-time. Journal of Research in Science Teaching, 11(2), 81-94.

Rummel, R.J. (1970). Applied factor analysis. Evanston, IL: Northwestern University Press.

Rupley, W.H., & Blair, T.R. (1987). Assignment and supervision of reading seatwork: Looking in on 12 primary teachers. The Reading Teacher, 40(4), 391-393.

Sandefur, J.T. (1985). State assessment trends. Washington, DC: American Association of Colleges of Teacher Education.

Sanford, J.P , & Everston, C.M. (1980). Beginning the school year at a low SES junior high: Three case studies. Austin: the University of Texas, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 195 547).

Shulman, L.S. (1986). Paradigms and research programs in the study of teaching. In M.C. Whittrock (Ed.). Handbook of research on teaching: Third edition. New York: Macmillan.

Soar, R.S., Medley, D.M., & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. Phi Delta Kappan. 65(4), 239-246.

Stallings, J.A. (1976). How instructional processes relate to child outcomes in a national study of follow through. Journal of Teacher Education, 27(1), 43-47.

Statistical Package for the Social Sciences. (1986). Chicago: SPSS, Inc.

Texas Education Agency Teacher Orientation Manual. (1988). Austin: Texas Education Agency.

Texas Teacher Appraisal System. (1986). Texas Teacher Appraisal System Instrument. Austin: Texas Education Agency.

Thompson, B.L. (1975). Secondary school pupils' attitudes to school and teachers. Educational Research, 18, 62-66.

Tucker, N. (1979). Could do better: Students reporting on teachers. Where, 152, 264-266.

Tuckman, B.W., & Oliver, W.F. (1968). Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 59(4), 297-301.

Veldman, D.J., & Peck, R.F. (1967). The pupil observation survey: Teacher characteristics from the students' viewpoint. Austin: Research and Development Center for Teacher Education, University of Texas. (ERIC Reproduction Document Service, No. ED 055 980).

Waxman, H.C. & Duschl, R.A. (1987). Using student perception data to improve preservice teachers' instruction and classroom environment. In B.J. Fraser (Ed.), The Study of Learning Environments (pp. 72-79). Western Australia: Curtin University of Technology.